

Relatório da Bolsa de Iniciação Científica
Métricas Econofísicas para Predição de Eventos Anormais em
Criptomoedas e sua Predizibilidade em Modelos de Aprendizado de
Máquina e Redes Neurais
Instituto de Física
Universidade Federal do Rio Grande do Sul

Bolsista: Henrique Alexandre Boneto Orientador: Sebastián Gonçalves

Porto Alegre, 11 de Maio, 2021

Introdução

Projeto

Esse projeto teve seu início em setembro de 2020 com o intuito de estudar as propriedades estatísticas das criptomoedas e a predizibilidade que métricas de econofísicas e de sistemas complexos têm quando usadas com modelos de aprendizado de máquina e redes neurais para prever eventos anormais (caracterizados como crashes e bolhas) desse mercado.

Criptomoedas

Criptomoedas têm se tornado um assunto de interesse para pesquisadores desde a apresentação do conceito de um novo sistema de pagamentos por [1] Satoshi Nakamoto em Bitcoin: A Peer-to-Peer Electronic.

Em janeiro de 2021, já existiam mais de quatro mil criptomoedas, todavia, apesar da maior parte delas não possuir um alto volume financeiro, as principais têm tido grande destaque. Os motivos para isso são diversos, mas deve-se principalmente pelas peculiaridades desses ativos, como por exemplo, um alto risco causado por especulações financeiras, ameaças de regulação e influência no mercado negro, o que gera dúvidas diante dos diversos entes da sociedade.

Econofísica

A econofísica é a área de estudo multidisciplinar que visa utilizar conceitos vindos da física para descrever eventos econômicos e melhor entender suas propriedades de um ponto de vista quantitativo e com explicação teórica fundamentada em princípios conhecidos.

Resumo

Na primeira etapa do trabalho, foi realizada a pesquisa bibliográfica sobre medidas especificamente advindas da física e de sistemas complexos [2] [3] que pudessem ser usadas para prever eventos excepcionais em mercados financeiros.

Após isso, foi feita a coleta de 2 anos de dados históricos de preços diários das dez criptomoedas mais capitalizadas e então realizada a análise de algumas de suas propriedades.

Metodologia

Para a realização do trabalho, está sendo utilizada a linguagem Python por conta da grande disponibilidade de bibliotecas pela comunidade, sobretudo para fácil manipulação de dados. Abaixo está a lista dessas bibliotecas junto com suas principais funcionalidades:

- Pandas: manipulação de dados.
- Numpy: operações vetoriais e algébricas.
- Scipy: computação científica.
- Matplotlib e Seaborn: utilizadas para criação de gráficos.
- python-binance: biblioteca de facilitação para chamadas na API (application programming interface) da corretora Binance, utilizada para busca de dados históricos de criptomoedas.

A Binance, atualmente, é considerada uma das maiores corretoras do mundo especializadas em criptomoedas. Está sendo utilizada especialmente por conter dados históricos em diferentes escalas de tempo, por exemplo, dados históricos de preços diários e intradiários, como de hora a hora ou mesmo minuto a minuto.

Dados

Os dados coletados nessa primeira fase do projeto, foram dados históricos dos preços do tipo OHLCV com escala diária em um período de dois anos, de 9 de Maio de 2019 até 8 de maio de 2021, contendo um total de 731 dias.

O acrônimo OHLCV significa Open High Low Close e Volume, sendo cada um desses representados em uma coluna da série temporal, além disso, também temos colunas representando as datas de abertura e fechamento de período. A tabela abaixo explica cada uma das colunas da série:

Variáveis dos dados e suas Descrições

Nome da Variável	Descrição
Open Time	Data de abertura do período
Close Time	Data de fechamento do período
Open	Preço de abertura do período
High	Preço máximo do período
Low	Preço mínimo do período
Close	Preço de fechamento do período
Volume	Volume financeiro negociado do período

Para esse trabalho foram coletados dados das dez mais capitalizadas criptomoedas no dia 08 de Maio de 2021 de acordo com a CoinMarketCap. Abaixo elas são listadas com nome e símbolo, respectivamente, em ordem decrescente de capitalização:

1. Bitcoin: BTC
2. Ethereum: ETH
3. Binance Coin: BNB
4. Ripple: XRP
5. Cardano: ADA
6. Litecoin: LTC
7. Chainlink: LINK
8. Ethereum Classic: ETC
9. THETA: THETA

10. Tron: TRX

Todas as criptomoedas diferentes do Bitcoin são denominadas Altcoins ou Alternative Coins (Moedas Alternativas), cada uma delas foi criada para um determinado propósito que aqui não será explicado.

Os dados históricos de preço coletados no trabalho, foram todos em relação ao Dólar Tether (com símbolo "USDT"), essa criptomoeda é denominada StableCoin, pois tem seu valor lastreado em outro ativo, nesse caso, é lastreada ao dólar americano e, portanto, 1 USDT sempre terá o mesmo valor que US\$ 1.

Abaixo, temos um exemplo de cinco dias de dados históricos do Bitcoin em relação ao Dólar Tether :

Dados Históricos OHLCV Diários para o Bitcoin

Open Time	Close Time	Open	High	Low	Close	Volume
2019-05-19	2019-05-19 23:59:59.999	7257.32	8275.09	7243.08	8148.48	65577.442058
2019-05-20	2019-05-20 23:59:59.999	8147.94	8156.03	7553.00	7938.15	65859.208564
2019-05-21	2019-05-21 23:59:59.999	7937.16	8042.32	7771.00	7904.87	52301.752247
2019-05-22	2019-05-22 23:59:59.999	7904.48	8016.00	7465.00	7628.43	49136.994589
2019-05-23	2019-05-23 23:59:59.999	7627.80	7940.98	7461.00	7851.51	49648.184701

Primeira Análise e Resultados

Começaremos analisando os preços de fechamento das criptomoedas que foram escolhidas. O preço de fechamento geralmente é utilizado para os cálculos por ser considerado o preço de consenso entre todos os participantes do mercado naquele período de tempo. A série temporal dos preços de fechamento aqui será denominada por

$$P = \{p_1, p_2, \dots, p_N\}$$

onde N é o tamanho total da série de dados, nesse caso N=731 dias.

A Figura 1, mostrada abaixo, contém os preços de fechamento diário. Através dela podemos perceber que houve um aumento acentuado dos preços aproximadamente a partir do dia 600 desde 8 de maio de 2019.

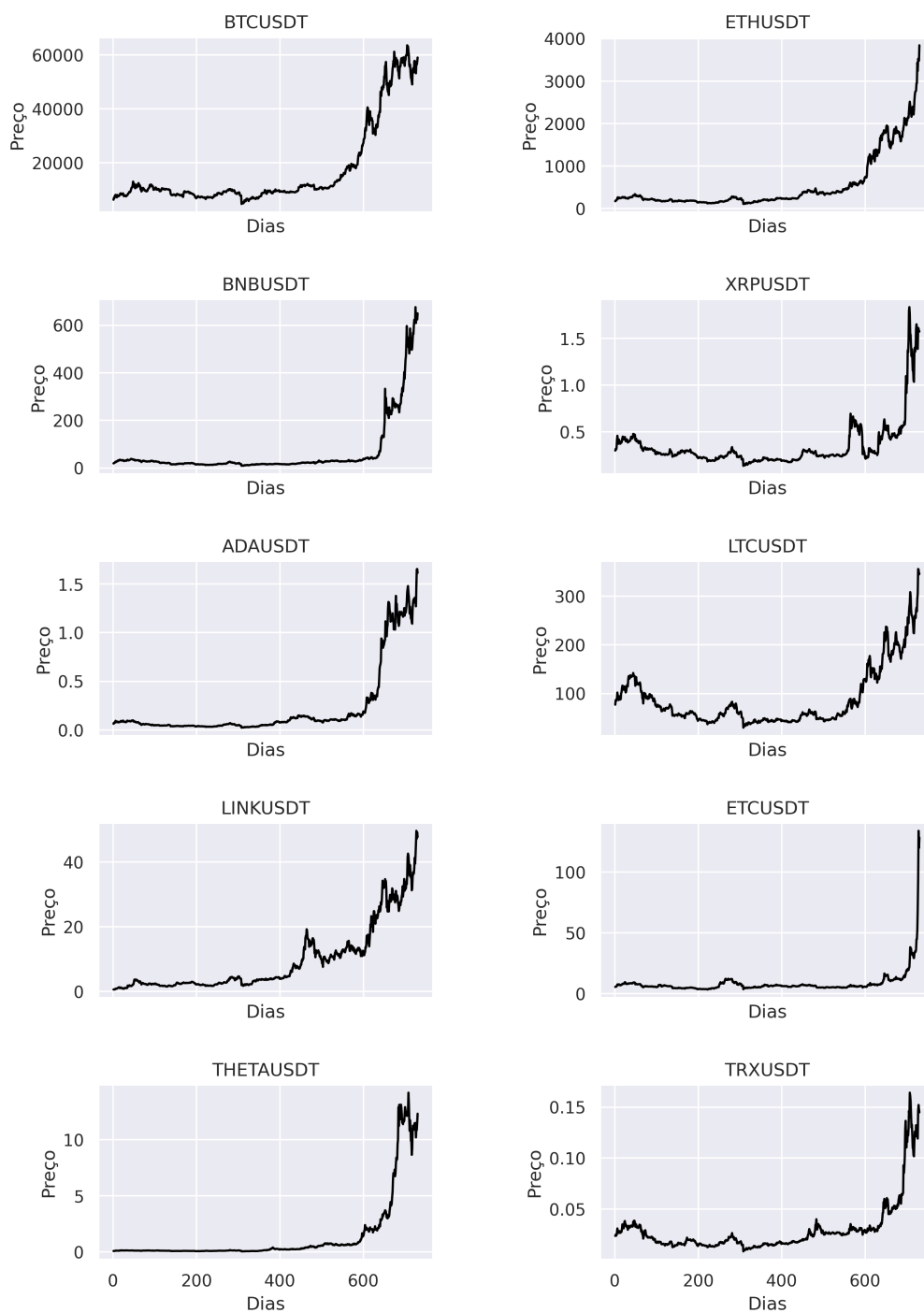


Figura 1: Preços de Fechamento para as Principais Criptomoedas

Teste de Estacionariedade

Foi realizado um teste de raiz unitária conhecido como Teste Aumentado de Dickey-Fuller, através dele podemos checar se a nossas séries temporais são estacionárias ou não. A hipótese nula H_0 do teste é de que os dados possuem raiz unitária e portanto não é estacionário, caso contrário, na rejeição de H_0 , concluímos que os dados são estacionários.

O teste foi feito para os preços de cada criptomoeda e obtidos o **valor-p**, neste caso, utilizando-se um nível de significância de 5%. Assim, caso o valor-p seja menor ou igual a 0.05 ($p \leq 0.05$), podemos rejeitar a hipótese nula e considerar a série estacionária. Abaixo temos a tabela com os valores-p de cada criptomoeda:

Teste de Estacionariedade Das Séries de Preços

Criptomoeda	Valor-p	Estacionário
BTCUSDT	0.9984	Não
ETHUSDT	1.0000	Não
BNBUSDT	1.0000	Não
XRPUSDT	0.9974	Não
ADAUSDT	0.9991	Não
LTCUSDT	1.0000	Não
LINKUSDT	1.0000	Não
ETCUSDT	1.0000	Não
THETAUSDT	0.9952	Não
TRXUSDT	1.0000	Não

Para transformar dados financeiros em estacionários, toma-se a primeira diferenciação da série. Os retornos financeiros de um período de $t - 1$ até t são definidos como:

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}}$$

Também temos os retornos logarítmicos, que são:

$$r_t = \ln(p_t) - \ln(p_{t-1})$$

Neste trabalho serão utilizados os retornos logarítmicos, pois são a primeira diferença dos preços logarítmicos. Eles são demonstrados na figura a seguir:

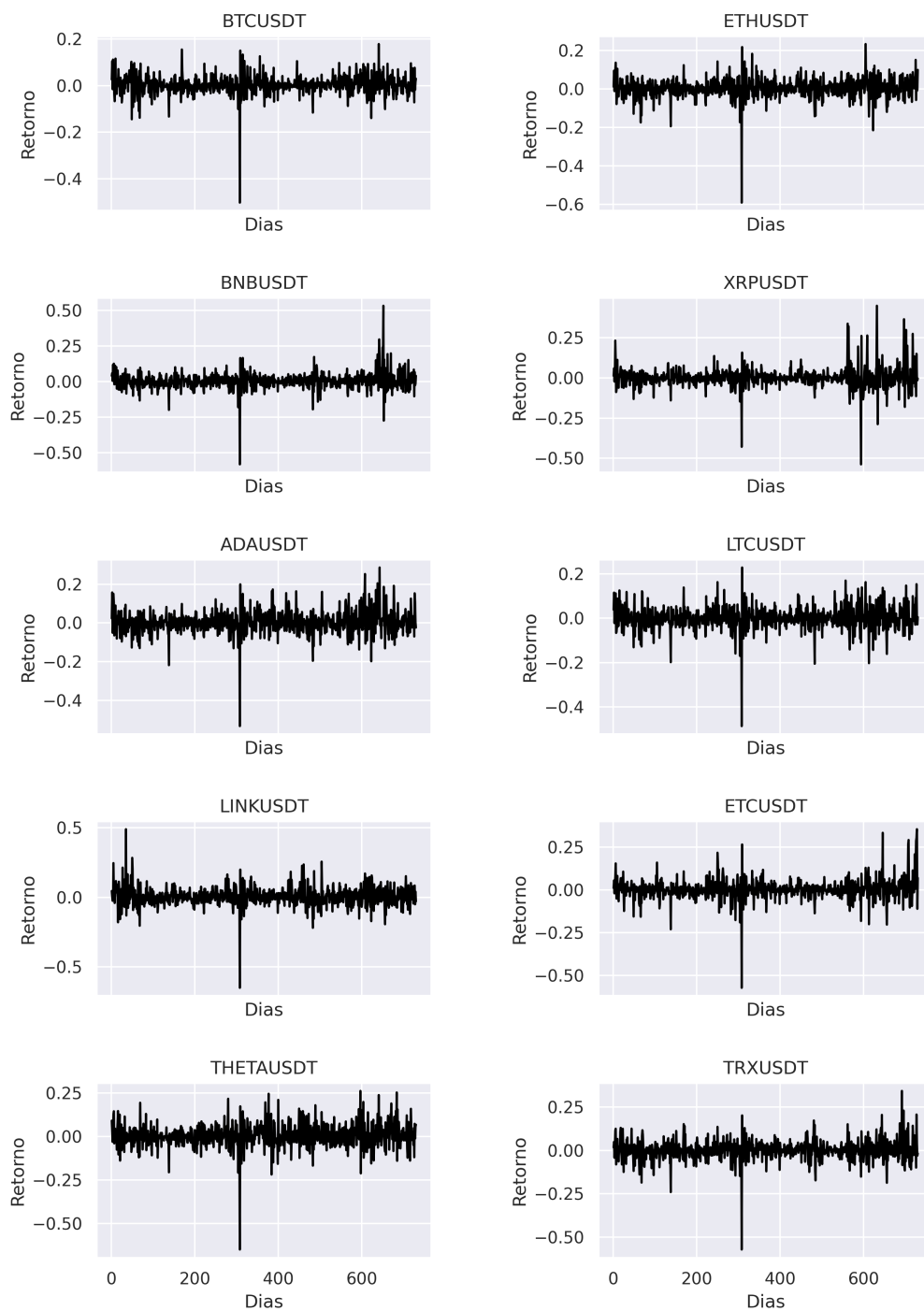


Figura 2: Retornos logarítmicos das principais criptomoedas

Resumos Estatísticos

Foram obtidos as seguintes estatísticas para os retornos das criptomoedas:

Estatísticas Descritivas dos Retornos Logarítmicos de diferentes Criptomoedas

Criptomoeda	Média	Desvio Padrão	Mínimo	Mediana	Máximo
BTCUSDT	0.003085	0.042047	-0.502607	0.001780	0.178449
ETHUSDT	0.004264	0.052305	-0.590534	0.003243	0.233501
BNBUSDT	0.004850	0.059414	-0.582287	0.002782	0.532404
XRPUSDT	0.002294	0.064753	-0.538657	0.001120	0.450100
ADAUSDT	0.004471	0.060040	-0.533075	0.004701	0.286441
LTCUSDT	0.002112	0.053353	-0.486668	0.000747	0.229360
LINKUSDT	0.005905	0.070391	-0.649747	0.003920	0.489638
ETCUSDT	0.004294	0.061629	-0.573024	0.002794	0.353934
THETAUSDT	0.006919	0.069709	-0.649636	0.005121	0.262890
TRXUSDT	0.002511	0.060020	-0.571268	0.002390	0.342818

Abaixo temos as figuras com as médias e os desvios padrões:

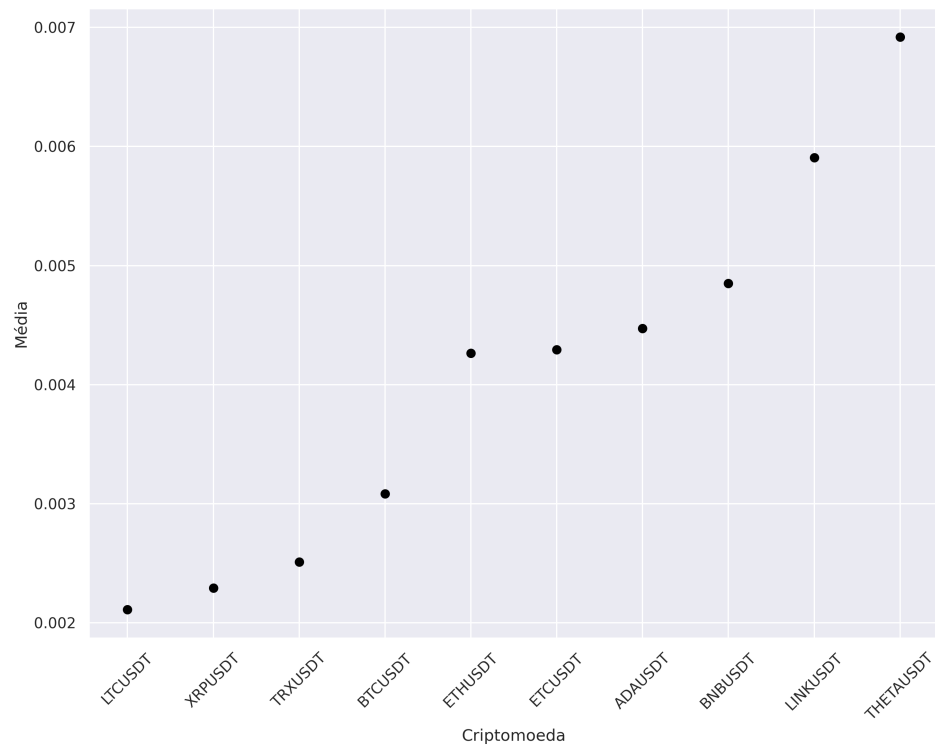


Figura 3: Médias dos retornos em ordem ascendente

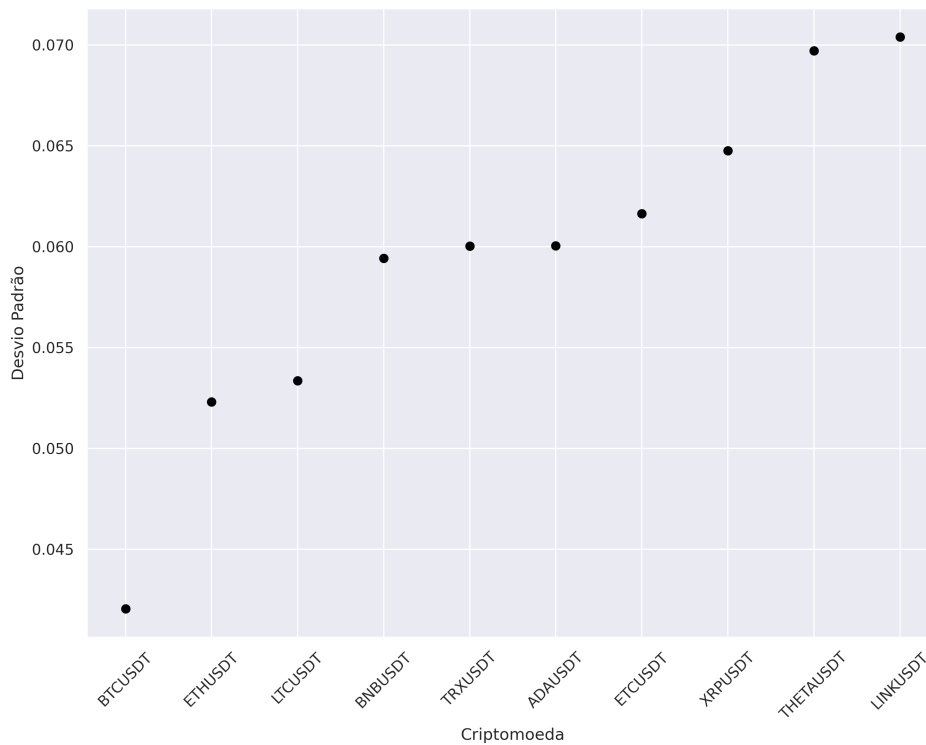


Figura 4: Desvio padrão dos retornos em ordem ascendente

Podemos notar que o LTC possui a média mais centrada em zero em comparação com as outras criptomoedas. O menor desvio padrão corresponde ao do Bitcoin, sendo assim considerada, a de menor volatilidade.

Distribuição dos Retornos

Foi obtido o gráfico dos retornos em forma de densidade (normalizando as contagens dos histogramas de forma que a sua área seja igual a 1), os dados foram divididos em 50 áreas igualmente espaçadas. A linha em cor preta é uma aproximação da distribuição utilizando-se de Estimativa de Densidade Kernel, com kernel gaussiano (KDE), já implementada dentro da biblioteca Seaborn. A linha em vermelho representa o fit (ajuste de curva) de uma distribuição gaussiana utilizando-se das medidas calculadas anteriormente (média e desvio padrão).

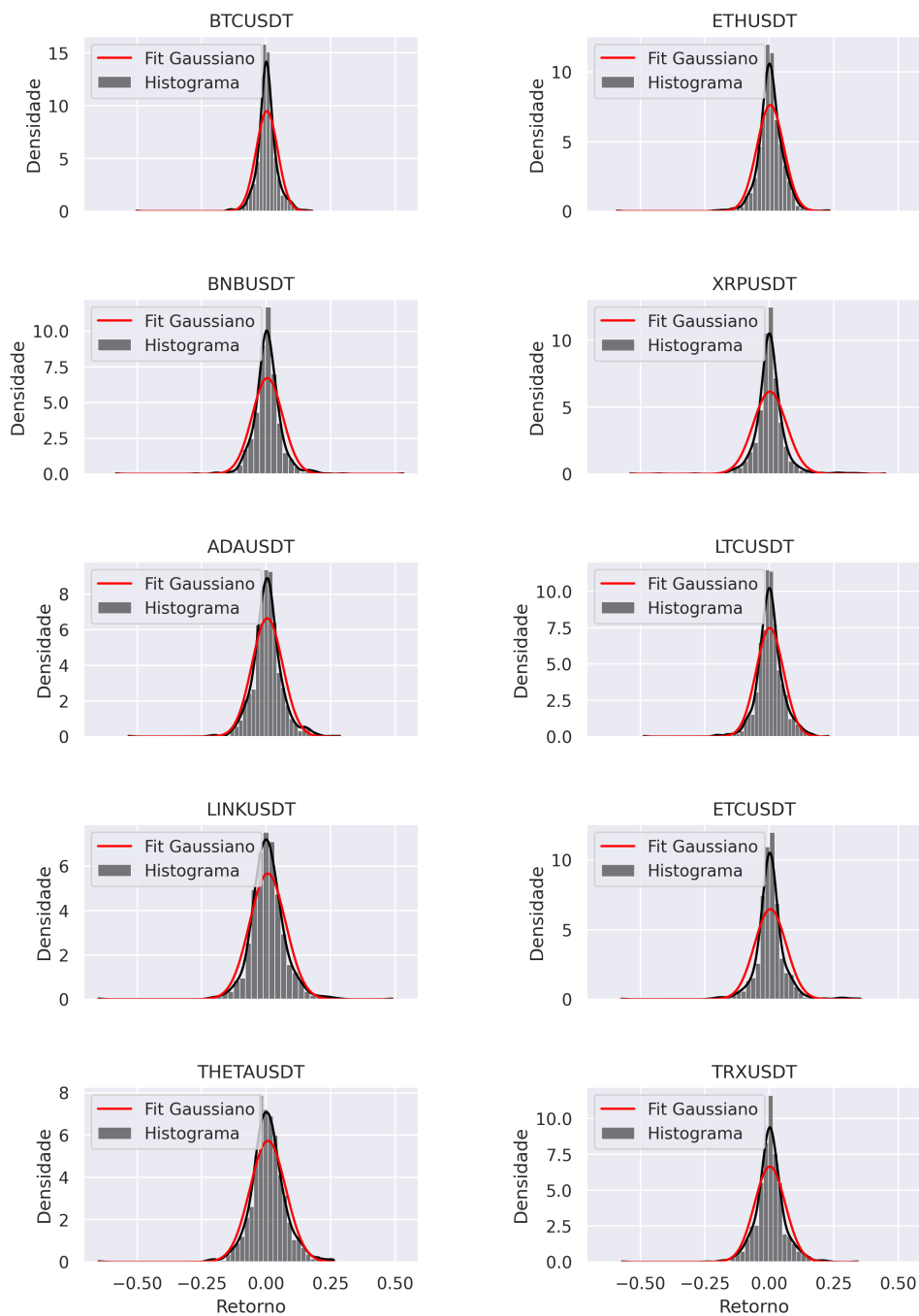


Figura 5: Distribuição dos retornos e fit gaussiano

Matriz de Correlação

A matriz de correlação é uma matriz \mathbf{C} de tamanho $M \times M$ onde cada elemento C_{ij} é a correlação de escolha entre duas variáveis. Foram obtidas as correlações dos retornos entre cada uma das criptomoedas, sendo o tipo de correlação calculado através do coeficiente de Pearson ρ , com $\rho \in [-1, 1]$.

Desse modo, um coeficiente Pearson positivo e mais próximo de 1 indica que se o retorno da criptomoeda X foi positivo, o retorno da criptomoeda Y tende a ser positivo. Por outro lado, se o coeficiente for negativo e mais próximo de -1, indicaria que caso a criptomoeda X tenha um retorno positivo, o retorno da criptomoeda Y tende a ser negativo.

A figura da matriz de correlação obtida encontra-se abaixo. Podemos perceber que todas as criptomoedas possuem correlação positiva entre seus retornos, quase todas com valor $\rho > 0.50$. Os maiores valores de correlação são encontrados no Bitcoin, sendo o maior com o Ethereum ($\rho = 0.82$). Esse fato pode indicar que o Bitcoin, que é a mais capitalizada das criptomoedas, "comanda" o mercado, pois as demais criptomoedas tendem a seguir a sua performance.

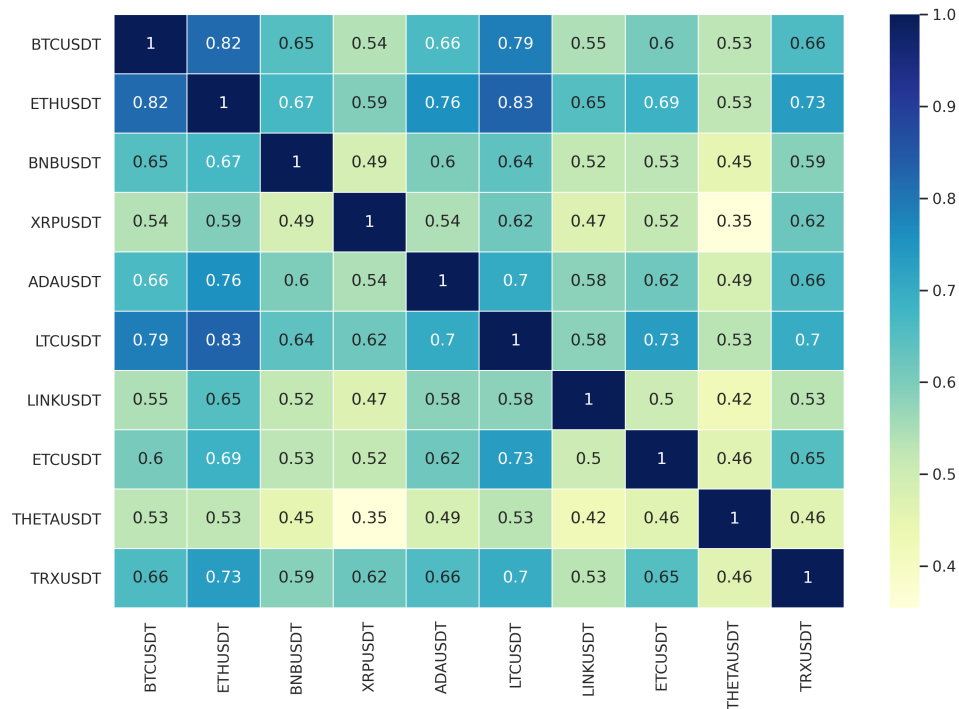


Figura 6: Matriz de correlação dos retornos entre as diferentes criptomoedas

Discussão

Como visto, as criptomoedas são altamente correlacionadas entre si e principalmente com o Bitcoin. Tal efeito pode ser perigoso, já que quedas e altas do Bitcoin podem movimentar grande parte do mercado. Assim, o estudo de métricas que possam prever eventos anômalos (crashes e bolhas) se faz necessário. Como exemplo disso, existem [2] trabalhos que utilizam métricas da física e de sistemas complexos para prever tais eventos, porém, ainda precisam da análise humana. Para contornar, poderia ser avaliado se tais medidas, unidas com modelos de aprendizado de máquina e redes neurais são eficazes.

Referências

- [1] Nakamoto, Satoshi. (2009). Bitcoin: A Peer-to-Peer Electronic Cash System
- [2] Soloviev V.N., Belinskiy A. (2019) Complex Systems Theory and Crashes of Cryptocurrency Market. In: Ermolayev V., Suárez-Figueroa M., Yakovyna V., Mayr H., Nikitchenko M., Spivakovsky A.

(eds) Information and Communication Technologies in Education, Research, and Industrial Applications. ICTERI 2018. Communications in Computer and Information Science, vol 1007. Springer, Cham. https://doi.org/10.1007/978-3-030-13929-2_14

[3] Bandt C, Pompe B. Permutation entropy: a natural complexity measure for time series. Phys Rev Lett. 2002 Apr 29;88(17):174102. doi: 10.1103/PhysRevLett.88.174102. Epub 2002 Apr 11. PMID: 12005759.

[4] Lauri Nevasalmi, Forecasting multinomial stock returns using machine learning methods, The Journal of Finance and Data Science, Volume 6, 2020, Pages 86-106, ISSN 2405-9188, <https://doi.org/10.1016/j.jfds.2020.09.001>.